

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

H04N 5/232

H04N 7/15 G01S 3/786

[12] 发明专利申请公开说明书

[21] 申请号 99805338.4

[43] 公开日 2001 年 6 月 6 日

[11] 公开号 CN 1298601A

[22] 申请日 1999.12.14 [21] 申请号 99805338.4

[30] 优先权

[32] 1998.12.22 [33] US [31] 09/218,554

[86] 国际申请 PCT/EP99/10066 1999.12.14

[87] 国际公布 WO00/38414 英 2000.6.29

[85] 进入国家阶段日期 2000.10.23

[71] 申请人 皇家飞利浦电子有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 E·科亨-索拉尔 A·P·马特尔

S·森古普塔 H·斯特鲁贝

J·卡维德斯 M·阿布德尔-莫塔勒布

A·埃尔加马尔

[74] 专利代理机构 中国专利代理(香港)有限公司

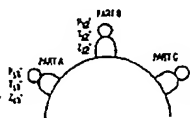
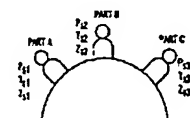
代理人 王 岳 王忠忠

权利要求书 5 页 说明书 11 页 附图页数 13 页

[54] 发明名称 自动确定对应于电视会议参加者的预置位置

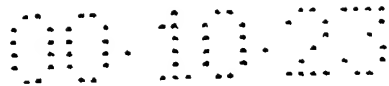
[57] 摘要

一种电视会议系统和方法,自动地确定对应于参加电视会议的参加者的适当的预置摄像参数。摄像机对电视会议空间取远景或摄全景并根据参加者的脸部寻找参加者。当一个参加者被检测到时,在参加者的中心位于摄像机视域中心期间,对于该参加者的预置摄像参数被计算。这个过程对于房间中的所有参加者继续进行。根据文化背景确定对每个参加者的最佳位置和相应的摄像参数。通过使摄像机变焦移出或对房间摄全景可以周期性地对预置值刷新。多台摄像机可用于连续地刷新预置值。



BEST AVAILABLE COPY

ISSN 1008-4274



权 利 要 求 书

1. 一种在电视会议系统中用于计算对应于参加者 (PartA, PartB, PartC) 的摄象机参数预置值的方法, 所述的方法包括:

- 提供一种具有倾斜、全景和变焦参数 (50) 的摄象机;

5 - 根据所述的电视会议系统的布局规定一个空间;

执行以下操作之一:

通过所有合适的摄全景值移动所述的摄象机, 所述的合适的摄全景值是由所述的电视会议系统所在的空间规定的, 和

10 使所述的摄象机取远景, 使得所有可能的参加者可被所述的摄象机观看到, 并使得在所述的空间中每个参加者的位置可被确定;

- 检测在所述空间内所述的参加者; 和

15 - 计算对应于所述的参加者的所述的预置值, 所述的预置值规定摄象机的视域, 所述的预置值是基于在所述的摄象机视域中所述的参加者的至少一个最佳位置, 将所述的参加者的头部中心与所述的摄象机视域的中心对准, 和将一个参加者的中心与所述的摄象机视域所述的中心对准。

2. 如权利要求 1 的方法, 进一步包括通过将标记与每个所述的参加者相关联来跟踪所述的参加者。

20 3. 如权利要求 1 的方法, 进一步包括刷新所述的预置值, 当预置值由使用者选取时, 所述的电视会议系统执行至少一项对预置值的调节, 当对应于预置值的参加者离开所述的空间时删除预置值, 并重复所述的实施步骤。

25 4. 如权利要求 1 的方法, 其中在所述的计算步骤中, 当多于一个参加者在所述的摄象机视域内时, 参加者被组合成为一个组合图象, 组合图象的中心被用于确定所述的预置值。

5. 如权利要求 1 的方法, 其中所述的检测步骤包括:

- 提供由多个象素组成的数字图象 (52);

- 通过检测肤色象素从数字图象产生二进制图象 (54);

30 - 在所述的二进制图象的亮度组分中排除对应于边线的象素, 从而产生二进制图象组分 (56);

- 将所述的二进制图象组分映射到至少一张图形中 (512); 和

- 将所述的被映射的二进制图象组分分类为脸部和非脸部类型,

其中脸部类型用作脸部的候选者 (514)。

6. 如权利要求 5 的方法, 进一步包括应用一种启发式方法的步骤, 所述的启发式方法包括以下步骤:

- 对每个所述的脸部候选者施加一种形态接近操作以产生至少一个接近的脸部候选者;
- 在所述的接近的脸部候选者中确定高方差象素;
- 在所述的接近的脸部候选者中确定所述的高方差象素和象素总数之间的比; 和
- 将所述的比与一个阈值作比较。

10 7. 如权利要求 5 的方法, 其中所述的排除步骤包括:

- 将一个障板加到包括一个被查看的象素的多个象素上;
- 确定在所述的被查看的象素和在所述的障板内放置的象素之间的方差; 和
- 将所述的方差与方差阈值作比较。

15 8. 如权利要求 7 的方法, 其中:

- 为了降低方差阈值, 所述的排除步骤被重复, 直到所述的二进制图象组分的大小低于组分大小的阈值为止; 和
- 在每次排除步骤以后实施对所述的组分的分类步骤。

9. 如权利要求 5 的方法, 其中所述的二进制图象组分被连接。

20 10. 如权利要求 5 的方法, 其中所述的分类步骤包括形成围绕所述的组分中一个被分类的组分的边界框并执行至少一项以下的操作:

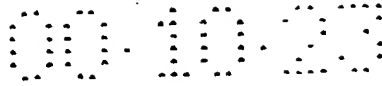
- 围绕所述的组分中一个被分类的组分形成一个边界框;
- 将边界框的面积与边界框阈值作比较;
- 将边界框的纵横比与纵横比阈值作比较;

25 - 确定面积比, 所述的面积比是在所述的被分类组分的面积和所述的边界框面积之间的比较, 和所述的面积比与面积比阈值的比较;

- 确定在所述的边界框内被延长物体的方向; 和
- 确定所述的边界框的中心和所述的被分类组分的中心之间的距离。

30 11. 如权利要求 5 的方法, 其中所述的映射步骤包括以下步骤:

- 将每个组分表示为一个顶点;
- 当在空间上接近和颜色上类似时, 将这些顶点用边线连接, 从



而构成所述的至少一个图形。

12. 如权利要求 11 的方法, 其中每条边线具有有关的权重, 并进一步包括以下步骤:

- 推断出每张图的最小跨度树;

5 - 将每张图相应的二进制图象组分分类为脸部或非脸部;

- 在每张具有最大权重的图形中除去边线从而形成两张较小的图形; 和

- 对每张所述的较小的图形重复所述的将相应的二进制图象组分分类的步骤, 直到围绕所述的较小图形的边界框小于图形阈值为止。

10 13. 如权利要求 1 的方法, 进一步包括: 提供至少一个第二摄象机, 用于通过执行所述的实施步骤刷新所述的预置值。

14. 一种电视会议系统, 包括:

- 至少一台具有全景、倾斜和变焦参数的摄象机 (50);

15 - 所述的参数具有分配给所述的电视会议系统的相应参加者 (PartA, PartB, PartC) 的预置值,

- 每个所述的预置值规定一种摄象机的视域, 并被确定于:

在由所述的电视会议系统规定的整个空间内进行将所述的摄象机摄全景和变焦操作中的一种,

检测一个参加者, 和

20 根据摄象机的位置规定预置值, 将所述的参加者放在最佳位置之一中, 所述的参加者的头部与所述的摄象机视域的中心对准的位置, 和所述的参加者的中心与所述的摄象机视域的中心对准的位置。

15. 如权利要求 14 的电视会议系统, 进一步包括通过将一个标记与每个所述的参加者相关联来跟踪所述的参加者的装置。

25 16. 如权利要求 14 的电视会议系统, 进一步包括用于刷新所述的预置值的装置, 通过当预置值由用户选取时, 所述的电视会议系统执行至少一项调节预置值的操作, 当对应于预置值的参加者离开所述的空间时, 删除预置值, 在整个所述的空间使所述的摄象机摄全景, 并在整个所述的空间使所述的摄象机变焦来实现。

30 17. 如权利要求 14 的电视会议系统, 其中当多于一个参加者在所述的摄象机视域内时, 参加者被组合成一个组合图象且组合图象的中心被用于确定所述的预置值。

18. 如权利要求 14 的电视会议系统, 其中所述的检测包括:

- 提供由多个象素组成的数字图象 (52);
- 通过检测肤色象素从数字图象产生二进制图象 (54);
- 在所述的二进制图象的亮度组分中删去对应于边线的象素, 从

5 而产生二进制图象组分 (56);

- 将所述的二进制图象组分映射到至少一张图形中 (512); 和
- 将所述的被映射的二进制图象组分分类为脸部和非脸部类型, 其中脸部类型用作脸部的候选者 (514).

19. 如权利要 14 的电视会议系统, 进一步包括至少一个第二摄象机用于刷新所述的预置值, 通过执行至少一项以下的操作来实现, 使所述的摄象机在整个所述的空间内摄全景, 并使所述的摄象机在整个所述的空间内变焦.

20. 一种电视会议系统, 包括:

- 至少一台具有全景、倾斜和变焦参数的摄象机 (50);
- 15 - 所述的参数具有分配给所述的电视会议系统的相应参加者的预置值, 所述的预置值规定摄象机的视域;

- 至少一台摄全景装置, 用于使所述的摄象机在由所述的电视会议系统规定的空间中摄全景, 和变焦装置, 用于使所述的摄象机取远景, 从而使所述的摄象机能观看由所述的电视会议系统规定的空间;

- 20 - 检测装置, 用于检测在所述空间中的参加者; 和

- 确定装置, 用于根据摄象机的位置确定所述的摄象机的预置值, 将所述的参加者之一放在最佳位置之一中, 所述的参加者的头部与所述的摄象机的视域中心对准的位置, 和所述的参加者的中心与所述的摄象机的视域中心对准的位置.

21. 如权利要求 20 的电视会议系统, 进一步包括用于跟踪所述的参加者的装置, 通过将一个标记与每个所述的参加者相关联来实现.

22. 如权利要求 20 的电视会议系统, 进一步包括用于刷新所述的预置值的装置, 通过当预置值由使用者选取时, 所述的电视会议系统执行至少一项调节预置值的操作, 当对应于预置值的参加者离开所述的空间时, 删除预置值, 使所述的摄象机在整个所述的空間摄全景, 和使所述的摄象机在整个所述的空間变焦来实现.

23. 如权利要求 20 的电视会议系统, 其中当多于一个参加者在所

述的摄象机视域内时，参加者被组合成一个组合图象，组合图象的中心被用于确定所述的预置值。

24. 如权利要求 20 的电视会议系统，其中所述的检测包括：

- 提供由多个象素组成的数字图象 (52)；

5 - 通过检测肤色象素从数字图象产生二进制图象 (54)；

- 在所述的二进制图象的亮度组分中除去对应于边线的象素，从而产生二进制图象组分 (56)；

- 将所述的二进制图象组分映射到至少一张图形中 (512)；和

- 将所述的被映射的二进制图象组分分类为脸部和非脸部类型，

10 其中脸部类型用作脸部候选者 (514)。

25. 如权利要求 20 的电视会议系统，进一步包括至少一台第二摄象机，用于刷新所述的预置值，通过执行至少一项以下操作来实现，使所述的摄象机在整个所述的空间摄全景，和使所述的摄象机在整个所述的空间变焦。

说明书

自动确定对应于电视会议

参加者的预置位置

5 本发明涉及电视会议技术的领域，尤其是涉及一种用于自动确定摄象机的适当的全景、倾斜和变焦参数的方法，这些参数相应于在电视会议设置中所希望的参加者的视域。

在电视会议期间，必须知道对每个参加者合适的摄象机参数，以便摄象机的视域可迅速地从一个参加者改变为另一个参加者。这些参数包括摄象机的适当的变焦、全景和倾斜—将被合起来称为摄象机“参
10 数”，这些与每个参加者有关的参数值是“预置”的。当会议进行时，用户需要有能力的快速观看不同的参加者；在短时间内频繁地从一个参加者改变为另一个。

现有技术设备需要用户人工地为每个电视会议的参加者设置摄象机参数。每个所用的摄象机被聚焦在一个参加者并启动一个初置开关。
15 例如，如果在会议中有三个人，开关 1 被用于代表对参加者 1 合适的摄象机参数；开关 2 用于参加者 2；且开关 3 用于参加者 3。当使用者希望在参加者 1 和 2 之间切换视域时，他只需要启动开关 2，摄象机被相应地移动和聚焦。然而，为每个参加者设置摄象机常常是一种冗长而乏味的过程，需要摄象机操作员或使用者付出时间。另外，每次一个参加者离开或进入房间，预置必须被相应地重新调节。如果一个参加者只是从他的原先位置移动，原先的摄象机预置将不再采用。很清楚，如果一个参加者在房间内从一个位置移到另一位置，这就是一个问题。然而，即使参数加在他自己的椅子内移动（即，向前，向后，
20 靠向一边，等），参数可改变，参加者可能不再处于焦点位置，不再处于摄象机视域的中心，或不再处于对于摄象机视域所希望的尺寸。

在美国专利 5,598,209 中，使用者可指出它希望观看的目标或人，系统自动地存储相对于目标中心的摄象机全景和倾斜参数。然而，所有在房间中的目标或人必须在使用者的控制下被确定地选取和存储，
25 这还是费时间。当一个参加者离开或进入房间时，也不提供对参数的更新。

自动确定预置位置的能力在会议布局中同样是有用的。一般，在

这些类型的房间中，摄象机预置是根据为每个个人使用的话筒。当一个参加者接通他的话筒时，使用涉及到话筒位置的摄象机预置。这是有问题的，因为如果话筒不工作，或者如果一个参加者的话筒被另一个谈话者使用，在谈话者和摄象机视域之间的适当相关将不发生。

- 5 因此，需要一种电视会谈系统，它自动地为所有参加者确定适当的摄象机参数，当参加者进入和离开房间时，也可自我调节。电视会议的目的是有效通信和交谈，如果一个使用者必须重新调节系统来使预置参数初始化或更新，那末这个目的被落空。在末端用户之间的交谈动态过程与一种产品（如在电视中）的动态过程不同。为了有利于
10 这种动态过程，希望使尽可能多的系统自动化，不求助于静态的取远景（zoomed out）得到意义不大的通信视域。

- 本发明的一个方面是一种在电视会议系统中计算对应于参加者的摄象机参数预置值的方法。本方法包括提供一种具有倾斜、全景和变焦参数的摄象机，并规定一个基于电视会议系统布局的空间。本方法
15 进一步包括通过所有适当的摄全景值移动摄象机，适当的摄全景值是由电视会议系统所在的空间规定的，摄取远景象机，使所有可能的参加者可被摄象机观察到，并使每个参加者在空间的位置可被确定。本方法进一步提供检测在空间内的参加者，并计算对应于参加者的预置值，预置值规定摄象机的视域，预置值是基于在摄象机视域中参加者
20 至少一个可选位置，将参加者头部的中心与摄象机视域的中心对齐，将参加者的中心位置与摄象机视域的中心对齐。

这一方面像以下的方面一样，允许自动检测和更新与电视会议中参加者对应的摄象机参加。

- 依据本发明的另一方面，一种电视会议系统包括至少一个具有全
25 景、倾斜，和变焦参数的摄象机。这些参数具有预置值，分配给电视会议系统的相应的参加者。每个预置值规定一种摄象机的视域，并确定于，在由电视会议系统所规定的空间中将摄象机摄全景和变焦操作中的一种，检测参加者，并根据摄象机位置规定预置值，将参加者放在最佳位置之一中，参加者的头部与摄象机视域的中心对齐的位置，
30 和参加者的中心位置与摄象机视域的中心对齐的位置。

依据本发明的另一方面，一种电视会议系统包括至少一个具有全景、倾斜和变焦参数的摄象机。这些参数具有预置值，分配给电视会

议系统的相应的参加者；预置值规定摄象机的视域。系统进一步包括至少一个摄全景装置，用于在由电视会议系统规定的空间中摇动摄象机摄全景，和变焦装置，用于使摄象机取远景，由此允许摄象机观看由电视会议系统限定的空间。一种检测装置被用于检测在此空间中的参加者。一种确定装置被用于根据摄象机的位置确定摄象机的预置值，将参加者之一放在最佳位置之一中，参加者的头部与所述的摄象机的视域中心对齐的位置，和参加者的中心部分与摄象机的视域中心对齐的位置。

10 本发明的一个目的是提供一种电视会议系统和方法，可自动地确定有关参加者适当的视域的摄象机参数预置值。

本发明的另一个目的是提供一种电视会议系统和方法，可以依据参加者数量和位置的变化连续地更新摄象机预置值。通过阅读以下的描述连同附图这些目的以及其他的目的将变得更清楚，其中同样的标号代表相同的部件。

15 图 1A, 1B 和 1C 分别是依据本发明的一个电视会议系统的房间、会议和桌子的布局图；

图 2A, 2B 和 2C 是示出在依据本发明的一个电视会议系统中当摄象机对一个房间摄全景时，一个参加者进入摄象机视域的图；

图 3 是在本发明中所用的一个摄象机的透视模型；

20 图 4 是示出带有各自指明的临时预置值的电视会议参加者的示意图；

图 5 示出一个参加者的中心部分偏离此参加者的摄象机视域中心的示意图；

25 图 6 示出电视会议的参加者带的各自被指明的已更新预置值的示意图；

图 7 是示出利用两台摄象机的本发明一种替代的实施方案的示意图；

图 8 是一种用于图象中象素的图形颜色的圆柱坐标系的示意图；

30 图 9 是三张表示 YUV 颜色域投影的图，指明皮肤的彩色象素所在的区域；

图 10A-10F 是原先的图象和各自的二进制图象，二进制图象是通过将基于彩色的象素分离形成的；

图 11 是用作说明一个 3×3 的掩膜是如何用作依据本发明的亮度变化检测的部件的示意图;

图 12A 和 12B 分别用作说明 4 和 8 类型的连通性的示意图;

图 13A 和 13B 示出在依据本发明的边缘被移动去后, 图 3C 和 3E 的图象看上去将像什么样的图象;

图 14 是示出施加到图象图 3F 上的边界框的例子的图象;

图 15 是示出依据本发明, 一个图象的组分是如何用各顶点来表示并被连接形成图形的一个系列图;

图 16A - 16D 是用作说明依据本发明的一种启发式方法的应用的一系列图象; 和

图 17 是一种详细说明包含在脸部检测中的一般步骤的流程图。

在图 1A 中, 示出一个电视会议系统, 其中参加者是围绕一张桌子就座的。图 1B 示出在大会型式安排中的参加者。一台摄象机 50 由控制器 52 控制, 从房间的一侧摇全景到另一侧, 很清楚, 摇全景的移动可在相同的地点开始和结束。例如, 如图 1c 中所示, 摄象机 50 可被安放在房间的中间, 参加者全都围着它就座。在这种类型的情况下, 摄象机 50 可完全按圆形旋转以便完全对整个房间摄全景, 在图 1B 中所示的大会方式安排中, 摄象机 50 可以采取多种摄全景的路径以覆盖不同的排。这些路径中每一条将具有不同的倾斜和可能还有不同的变焦 (如果参加者在离摄象机基本相同的半径距离上互相直接向上就座, 变焦可以是相同的)。又, 在大会方式安排中, 摄象机 50 可被放置在房间的中间, 然后摄全景的移动可能如图 1C 中所示的那样需要完全的旋转。

为简单起见, 图 1A 中所示的安排现在将作进一步的描述, 然而应该明白, 相同的思路将适用于所有提到的安排并且其他的安排对于本领域的普通技术人员也是明白的。本发明将由电视会议系统的可调节性所规定的任何空间服务。三位参加者 ($Part_A$, $Part_B$ 和 $Part_C$) 被示出, 但, 更多的参加者也可被包括在内。

当摄象机 50 从房间的一侧摇全景到另一侧时, 参加者将看上去是移动着通过摄象机的视域。正如在图 2A - 2C 中所示的那样, 一个参加者出现在摄象机视域的不同部分, 取决于摄象机摄全景的位置。也可从图中看出, 对于三个不同的摄全景的位置 (P_1 , P_2 , P_3), 倾斜



(T) 和变焦 (Z) 仍然是相同的。在初始摄象机浏览期间, 其他参数 (也就是倾斜或变焦) 之一可在适当的范围内移动, 同时其余的两个参数保持不变。另一种可能性是, 如果摄象机 50 已设置其变焦参数, 以致整个房间可马上看见 (假定足够的信息可被搜集以确定不动的参加者的位置, 以下将更清楚地讨论)。再次, 为了简单, 摄象机摄全景的思路将被描述, 但应该明白, 带有适当变化的其他建议可被实现, 这对于本领域的技术人员是清楚的。

在初始摄全景期间, 摄象机处理的每个帧被分析以确定是否一个参加者是在帧内。作这种确定的一种方法在以下的参加者检测节中详细说明。很清楚, 其他的方法可实施。对于每个被检测的参加者, 摄全景的摄象机将检测包括参加者的多个帧。例如, 如果一台摄象机对一个房间处理一千个帧, 如果一个参加者在每帧中都出现, 这可能被解释为一个参加者。

为了避免参加者实际数目增多的这个问题, 每个被检测的参加者被打上标记。对于每个被处理的帧计算每个被检测的参加者的质量中心。然后, 第二个, 包含潜在参加者的相继的帧与以前的第一帧作比较, 看看是否摄象机正在观看一个新的参加者或者只是包括相同参加者的另一帧。实现这种比较的一个方法是执行一种基于第一中心和摄象机已从第一位置移动的量的几何外推。如果第二帧包括与第一帧相同的参加者, 这将得到该中心应该所在的近似位置。同样, 可以计算第二帧的质量中心, 然后随同在第一帧被观看的位置和第二帧被观看的位置之间摄象机已知的移动与第一中心作比较。另外, 可为每个被检测的参加者建立署名, 然后在相继帧中参加者的署名可与起始的署名作比较。署名方法在技术上是已知的。某些署名技术的例子在以下的参加者识别和位置刷新章节中讨论。一旦确定一个参加者的图象被放置在一个帧内, 可计算临时预置值。

参考图 3, 示出一台摄象机的透视模型。摄象机的传感器 56 有一个主要点 PP, 其 x, y 座标分别为 PP_x 和 PP_y 。透镜 58 的中心位于离主要点 PP 的焦点长度 f 上。摄象机变焦的改变是通过改变焦点长度实现的。较短的 f 意味着宽的视域 (“取远景”)。在摄全景参数方面的变化是使传感器围绕全景轴有效地旋转。在倾斜参数方面的变化是使传感器围绕倾斜轴旋转。

当一个目标或参加者 62 进入摄像机的视域时, 如果包含参加者的两个帧可得到的话, 参加者在空间的位置可利用常规方法确定。这是因为主要点 PP (现为 60) 和焦点 f 的位置是已知的。当摄像机 50 对一个房间摄全景时, 获得包含参加者的多个帧, 所以每个参加者在空间的位置可被确定。如果摄像机正在变焦移出而不是摄全景, 可能需要两种性质不同的测量以确定位置。一旦一个参加者的位置已知, 可由处理器 54 (图 1A-1C) 计算临时预置值。

为了计算临时预置值, 如以上对参加者打标记那样, 利用已知技术确定参加者的中心。例如, 可以计算参加者的轮廓平均和其质量中心。然后将中心点放在摄像机视域的中心以产生, 例如, 图 1 中 Part_A 的预置值 Psa, Tsa, 和 Zsa。对于房间中的所有参加者重复这些摄全景和预置值计算过程, 从而, 也确定一开始在房间中有多少参加者。在会议的开始期间全部被实施, 往后在更新的程序期间可被重复, 以下要更充分地进行描述。

一旦如图 4 所示, 在房间中的所有参加者被打上标记, 所有的临时参数被计算出, 摄像机 50 对房间实施第二次摄全景 (取远景)。每个预置视域被进一步细调, 因为在初始摄全景阶段中实施的标定一般来说将是不够精确的。

如图 5 中所示, 摄像机视域的中心与每个参加者的头部的中心作比较。参数被调节, 使得在摄像机视域中这些中心调成一致。一旦预置值被精调, 计算与每个参加者的“最佳”视域对应的预置值。这可能是不同的, 与社会文化有关。例如, 一个参加者的头部和躯体可以取整个帧的 30-60% - 如在美国新闻节目中那样。最佳视域产生刷新的预置值 Psn', Tsn' 和 Zsn', 如图 6 中所示, 这些值被连续地刷新, 取决于系统是如何构成的和刷新是如何实施的, 以下要解释。如果摄像机对着一个参加者, 参加者移动时, 将计算新的最佳位置, 因此摄像机的预置值将被连续地调节。

摄像机可根据声频跟踪, 视频跟踪, 由使用者作选择, 或由技术上已知的任何其他技术作选择, 对参加者聚焦。单独的声频跟踪是有限度的, 因为当人进一步离开时在精度上被降低, 并且不可能由本身使用, 因为它一般有 4-5 度的误差, 当参加者停止谈话时不可能进行跟踪。

一旦参加者被检测到，每个参加者联系着一个名字，例如，图 1 的三个参加者可被识别为 A、B 和 C，以致使用者可仅指出他希望观看参加者 A，摄像机将移到对 A 最佳的最佳值。另外，系统可被编程监听关于每个参加者的某些特征，因而对参加者打上标记。例如，为每个参加者建立署名，个人的衬衫颜色，可取得的话音型式，或者脸部和

5 话音的组合可被用于构成与参加者有关的标记。利用这种附加的信息，如果参加者 A 绕着房间移动，系统将知道哪个参加者正在移动，将不会被参加者 A 通过对应于参加者 B 的参数的视域走动弄混淆。而且，如果两个参加者的位置互相靠得足够近，以致他们共享一个摄像机的视域，那末两个参加者可被认为是一个参加者，摄像机聚焦在他们的

10 图象组合的中心。

如上所述，本发明的一个好处是当房间的参加者的动态状况改变时，允许预置值能被自动地调节。很清楚，如果选取一个预置值，相应的参加者离开房间，系统将感知这点并刷新预置值。另一种刷新

15 的方法是每次一个新的预置值被选取时，在摄像机 50 移到所选的预置值以前，摄像机 50 将变焦移出（或摄房间全景）看看，是否任何人已经进入或离开房间。摄像机 50 可被周期性地控制，即使在被命令观看一个所选的参加者时，临时停止观看参加者，对房间摄全景或取远景，看看是否参加者的数目已经改变。另一种技术是认出一个参加者并不

20 在他应该呆的地方，例如，如摄像机 50 被告知从参加者 C 的预置值移到例如（图 1）参加者 A，如果参加者 B 已经离开房间，系统可认识到这点并作出适当的调节。还有另一种刷新技术包括摄像机 50 对整个房间摄全景（或者取远景），或者是周期性地或者是每次一个新的预置值被选取时进行。

25 参考图 7，示出第二实施方案。除了增加第二摄像机 64 以外，该实施方案示出与图 1A 中相同的特性。实施初始标定与上述的相同。然而，在会议期间，一台摄像机被用于对适当的参加者聚焦，而另一台被用于连续地刷新预置值，正在刷新的摄像机可被连续地取远景，使得它可确定何时一个参加者离开或进入房间。另一种方案是，正在刷

30 新的摄像机可连续地对房间摄全景，对预置值作适当的刷新。两台摄像机通过，例如，处理器 54 共享预置信息。很清楚，可使用多台摄像机。例如，一台摄像机可分配给计划出席会议的每个人，然后一台

附加的摄象机可被用作刷新该摄象机。

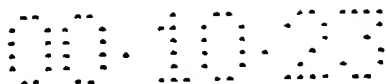
确定是否一个参加者位于摄象机视域内的一种方法是确定是否有一张脸位于摄象机观看到的图象内。在图象中的每个象素通常在 HSV (色调, 饱和, 值) 颜色域中表示。这些值被映射到圆柱坐标系上, 如图 8 所示, 其中 P 是值 (或亮度), θ 是色调, r 是饱和。由于圆柱坐标系的非线性, 其他的颜色空间被用于近似 HSV 空间。在本应用中, 采用 YUV 颜色空间, 因为大多数存储在磁介质上的视频材料和 MPEG2 标准都使用这种颜色空间。

将 RGB 图象变换到 YUV 域, 并进一步投影到 VU, VY 和 VU 平面, 产生如图 9 所示的图形。圆形段代表 HSV 域的近似值。当对应于皮肤颜色的象素画在 YUV 空间中时, 通常落在所示的这些圆形段内。例如, 当一个象素的亮度具有值在 0 和 200 之间, 色度 U 通常对于一个肤色的象素具有在 -100 和 0 之间的值。这些是基于实验的一般值。很清楚, 可为每个所用的摄象机实施一种颜色训练操作。然后训练的结果将被用于产生更精确的肤色段。

为了检测一张脸, 在图象中的每一象素被查看以便辨别是否是皮肤颜色。那些是皮肤颜色的象素被从图象的其余部分编组, 并因而被留作潜在的脸的候选者。如果一个象素的至少一个投影并未落在皮肤组段的边界内, 则该象素被认为不是皮肤颜色, 从作为潜在脸的候选者的考虑内除去。

由皮肤颜色检测构成的合成图象是二进制的, 因为它示出或者是图象的皮肤颜色部分, 或者不是皮肤颜色的部分, 如图 10B, 10D 和 10F 中所示, 对应于图 10A, 10C, 和 10E 中原先的图象。在图中, 白色表示皮肤颜色, 黑色表示非皮肤颜色。如图 10A 和 10B 中所示, 单单这个检测步骤当其中有脸部时, 可删除大部分的图象。利用颜色和形状的现有技术可为简单的背景工作, 如图 10A 所示。可是, 看在图 10C 和 10D, 图 10E 和图 10F 上, 很清楚仅用颜色和形状来检测对于检测脸部不可能是足够的。在图 10C-10F 中, 在背景中的物体如皮革, 木头, 衣服, 和头发, 具有与皮肤类似的顔色。正如从图 10D 和 10F 中所见, 这些皮肤颜色的物体与脸部皮肤紧邻地放置, 所以脸部本身难以检测。

在象素用颜色分离后, 位于边缘上的象素从考虑中排除。一个边缘在亮度等级方面从一个象素到下一个象素是变化的。通过取每个皮



肤颜色像素并计算在亮度组分中围绕它的像素中的方差实现这种排除；高方差指明是一个边缘，如图 11 中所示，一个或者是 3×3 或者是 5×5 像素尺寸的框（“窗”）被放在肤色像素的顶上。很清楚，可以采用方框以外的其他障板。

5 方差被定义为

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2$$

其中 μ_x 是被查看的窗中所有像素的平均值。“高”方差的水平将是不同的，取决于脸部和所用的摄象机。因此，采用一种迭代程序从非常高的方差水平开始往下工作到低的方差水平。

10 在方差迭代的每一步，如果在肤色像素周围的一个窗中的方差大于对该迭代测试到的方差阈值，这些像素从面部的考虑中排除。所有的像素在一次迭代中被查看以后，为面部的特征查看合成的已连接的组分，以下将更充分地描述。已连接的组分是具有相同的二进制值（对于面部颜色是白色）并被连接的像素。连通性可以是 4 或 8 类型的连通性。正如图 12A 中所示，对于 4 类型的连通性，中心像素被考虑“连接”仅直接邻近它的像素，在邻近框中用“1”表示。在 8 类型的连通性中，如图 12B 所示，对角线接触中心像素的像素也被考虑为“连接”此像素。

20 如上所述，在每次迭代以后，已连接的组分被在组分分类步骤中查看，看看是否它们可能是一张脸。这种查看包含看五个性质不同的准则，是根据围绕每个合成的已连接的组分所画的边界框进行的；它们的例子示于图 14 中，是以图 10E 的图象为基础的，这些准则是：

25 1. 边界框的面积与一个阈值相比较。这种方法辨认面孔，一个面孔通常将不是非常大或非常小。

2. 边界框的纵横比（高与宽比较）与一个阈值相比较。这种方法辨认人的面孔通常落在纵横比的一个范围内。

30 3. 检测到的肤色像素的面积与边界框面积之比与一个阈值相比较。此准则辨认面孔，由人的面孔覆盖的面积将落入边界框面积的百分数的范围内。

4. 边界框内被伸长物体的方向。有许多已知的方法确定一系列象素的方向。例如，可以确定医学轴，从此轴可找出方向。通常，面孔并不围绕此轴（“Z-轴”）较大地旋转，该轴与具有图象的平面垂直，所以带有相对于Z轴旋转的被伸长物体的组分被从考虑中排除。

- 5 5. 边界框的中心和被查看的组分的质量中心之间的距离。通常，面孔位于边界框的中心内，并例如，将不全部位于一侧。

对于方差的迭代由此继续分析图象进入较小组分，直到组分的大小低于一个阈值为止。图 10C 和 10E 的图象在方差迭代过程以后分别被示出转变为图 13A 和 13B。正如可被辨别的那样，作为方差迭代的结果，在图象中的面孔与背景中非面部肤色区分离。经常，这引起被检测到的肤色区被分裂，如图 13B 中举例的那样。这种情况的发生是因为存在阻塞部分面孔的物体（如眼镜或面部的头发）或者由于高的方差除去若干部分。因此利用合成的组分通过他们自己来寻找一个面孔将是困难的。在方差迭代和组分分类步骤以后仍然可能是脸的部分的组分被连接形成如图 15 中所示的图形。用这种方法，具有类似特性和空间上靠近的肤色组分被一起编组并接着进一步查看。

参考图 15，每个合成的组分（颜色检测，边缘排除，和组分分类步骤的幸存者）用图形的顶点表示。如果这些顶点在原先图象中在空间上是靠近的，和在原先的图象中具有类似的颜色，在把这些顶点连接。如果：

$$|Y_i - Y_j| < t_n \wedge |U_i - U_j| < t_n \text{ AND } |V_i - V_j| < t_n$$

两个组分 i 和 j 具有类似的颜色。

其中 Y_n , U_n 和 V_n 是第 n 个组分的亮度和色度的平均值， t_n 是阈值。阈值是基于面孔中 Y, U, 和 V 中的方差并保持足够高，以致相同面孔的组分将被认为是类似的。如果组分之间的距离小于阈值，就认为这些组分在空间上是靠近的。空间上的要求保证空间上远离的组分不在一起编组，因为一个面孔的部分通常并不位于图象中空间上远离的部分。

在顶点之间的连接被称为边线。每个边线给出与两个顶点之间的欧几里德距离成比例的权重。将顶点连在一起将得到一张图或一组拆散的图。对于每一张得到的图形，推断出最小跨距树。最小跨距树通

常被定义为一张图的子集，其中所有的顶点仍然是被连接的，该图的边线长度之和为尽可能的小（最小权重）。然后利用在上面提到的组分分类步骤中规定的外形参数，将对应于每张所得的图形的组分分类为脸部或非脸部，然后通过将最弱的边线（具有最大权重的边线）除去将每张图分成两张图，再次查看所得的图形对应的组分，继续划分，直到围绕合成的图形构成的边界框的面积小于阈值为止。

通过分析和查看对于一张脸的每张图，在一个图象中一组所有可能的脸部的位置和其尺寸被确定。该组可包含大量的虚假正片，所以一种启发式方法被用于除去某些虚假的正片。寻找所有的脸部特征（也就是鼻，嘴，等）将需要一种样板，这将得到太大的搜索空间。然而，实验已经表明，那些脸部特征具有带高方差的边线。通过查看一张潜在的面孔内的高方差象素与此潜在面孔中象素总数的比例，可以排除许多虚假的正片。

以上提到的启动式方法是首先通过对图象内的脸部候选者应用一种形态接近操作来实现的。正如在技术上已知的那样，选取一种障板并应用到一个潜在的脸部区域内的每个象素。例如，可采用一个 3×3 障板。一种膨胀算法被用于扩展脸部候选者组分的边界。然后，一种腐蚀算法被用于消除来自边界的象素。本领域的技术人员将认识到，按这种次序执行的这两种算法将充满组分之间的缝隙，并也将按基本相同的比例保持这些组分。很清楚，我们可以执行多次膨胀，然后是多次腐蚀步骤，只要两者被应用相等数量的次数。

现在，在脸部候选者区域内带有高方差邻居的象素与脸部候选者区域中的象素总数作比较。参考图 16A 到 16D，利用以上描述的方法，为潜在的脸部候选者查看图 16A 中一个原先的图象，以实现图 16B 中所示的二进制图象。对二进制图象执行形态接近操作，得到图 16C 中所示的图象。最后，位于图 16C 的图象中的带有高方差的象素被检测，如图 16D 中所示。然后可以确定高方差象素与象素总数之比。通过图 17 中所示的步骤 S2 - S16 概要说明整个参加者检测方法。

可以看出，通过控制摄像机观看由电视会议系统规定的空间，对应于参加者的摄像机参数预置值可被自动地计算和连续地刷新。

通过描述优选实施方案，应该明了，可以作各种改变而不偏离在所附的权利要求中被更清楚地规定的本发明的范围和精神。

00.10.23

说明书附图

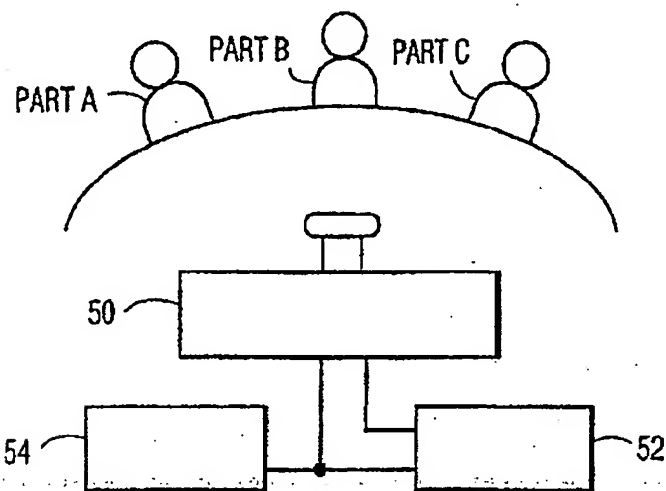


图 1A

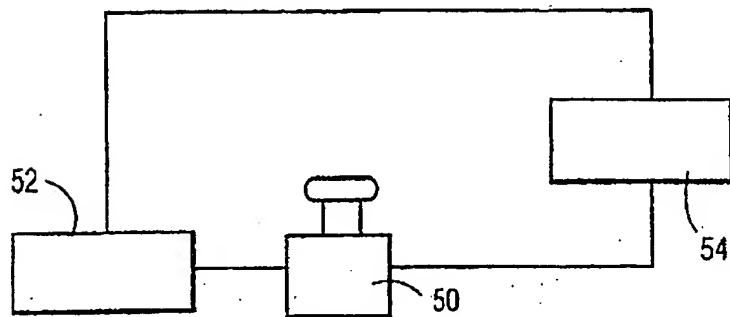
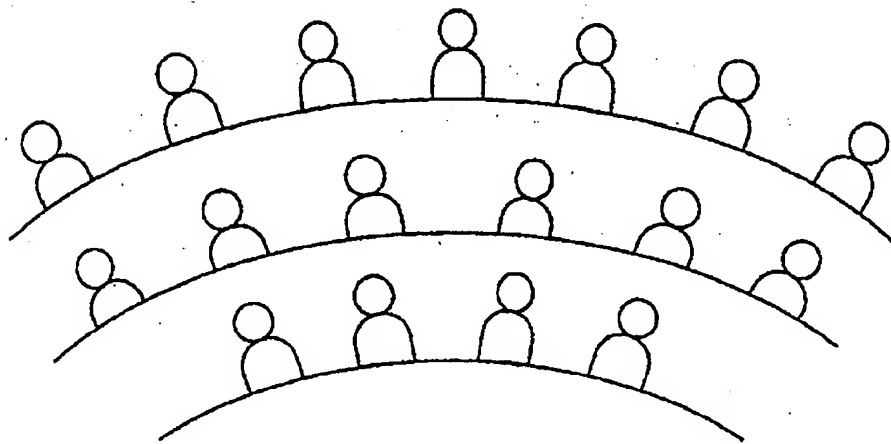


图 1B

00:10:23

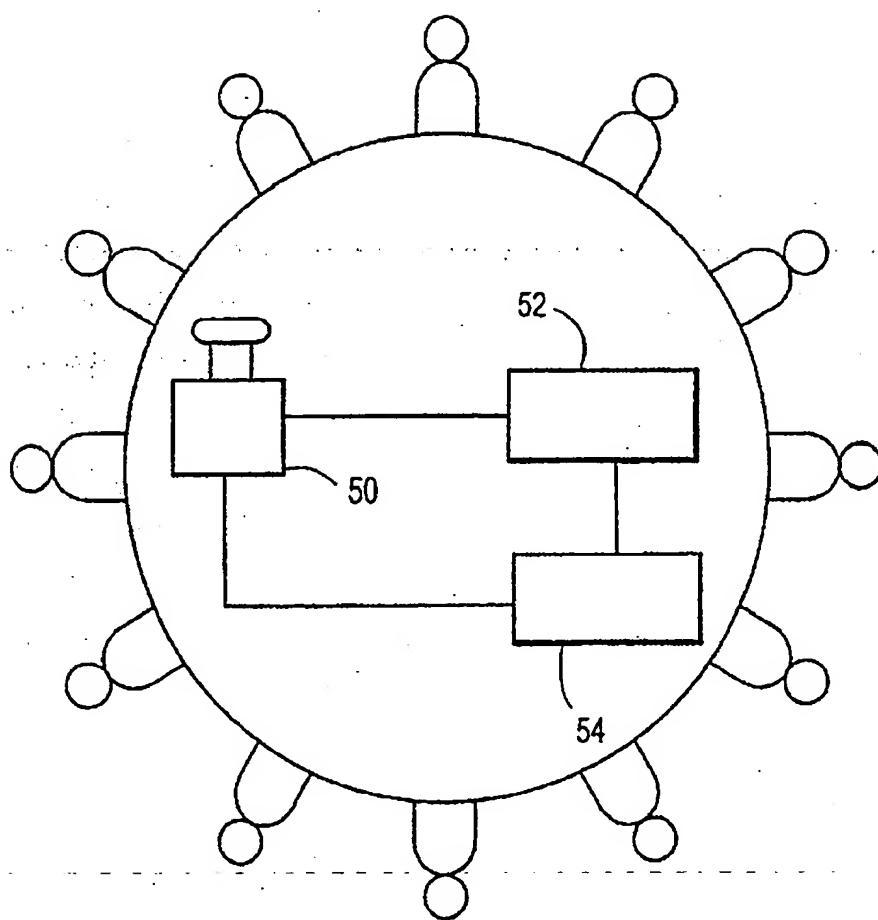


图 1C

00.10.23

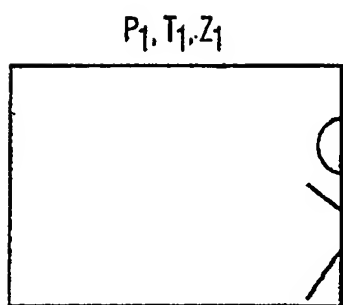


图 2A

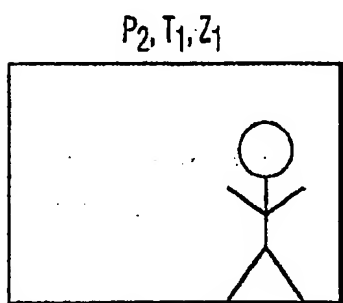


图 2B

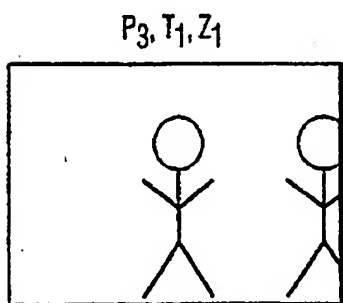


图 2C

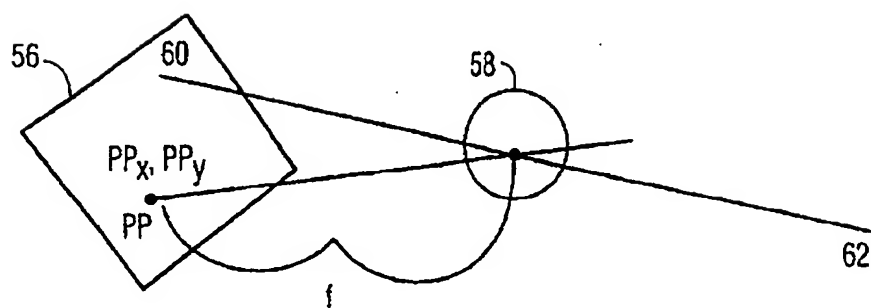


图 3

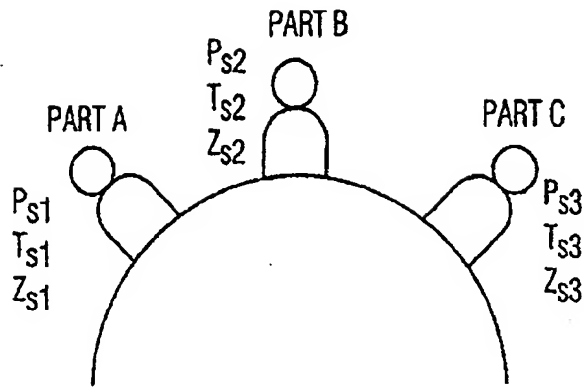


图 4

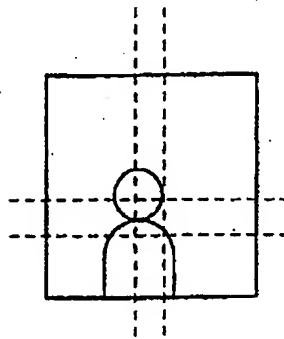


图 5

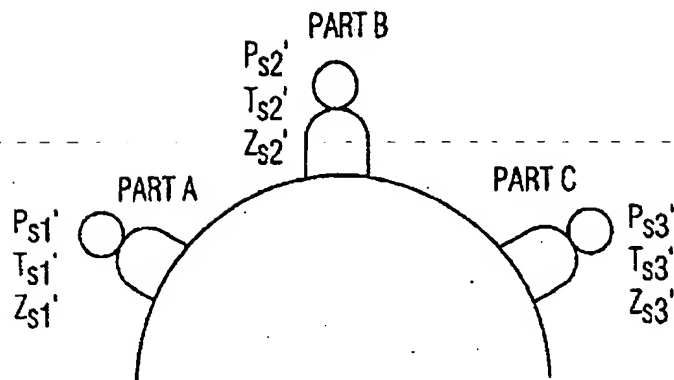


图 6

00-10-23

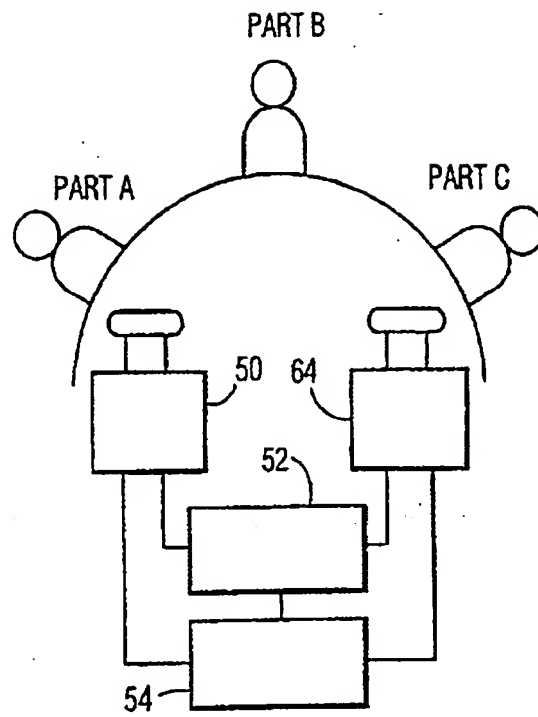


图 7

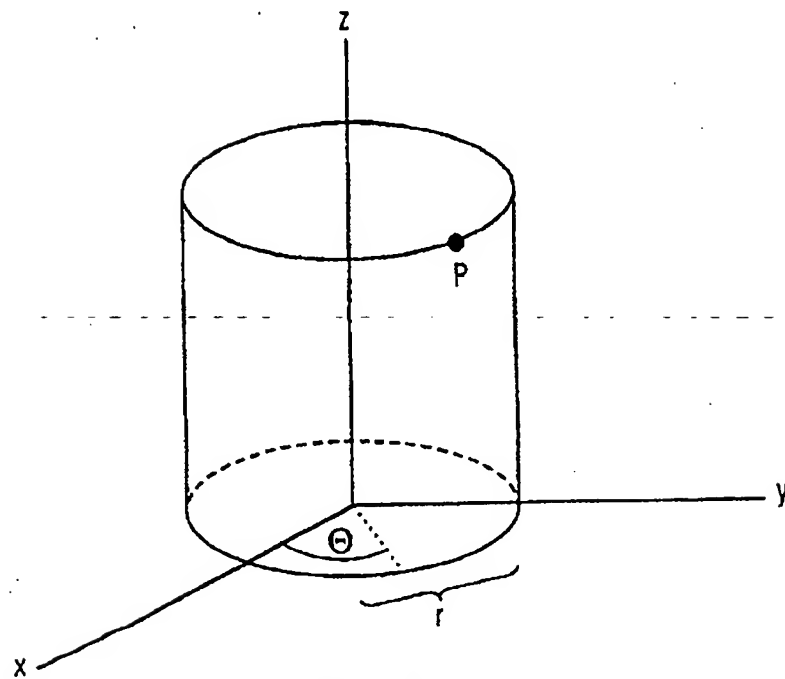


图 8

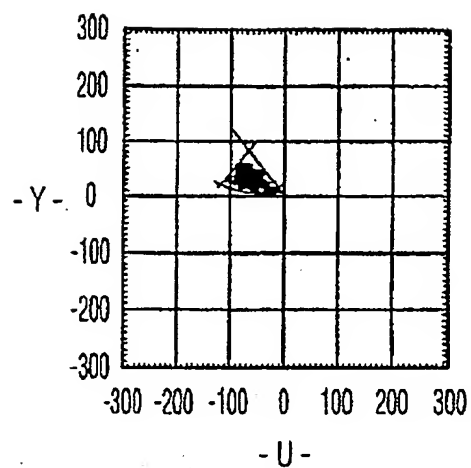


图 9A

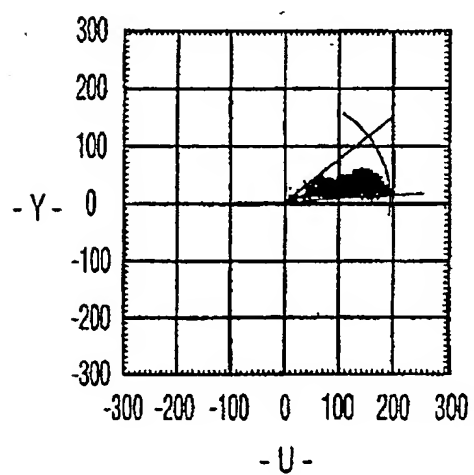


图 9B

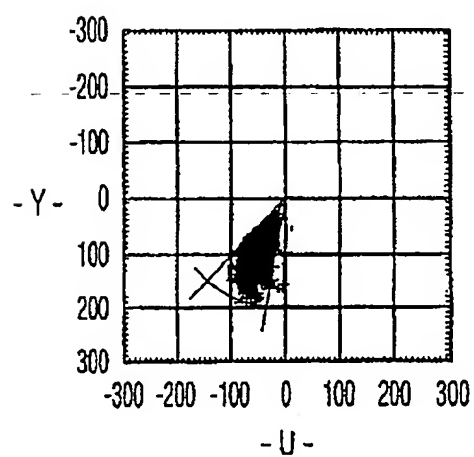


图 9C

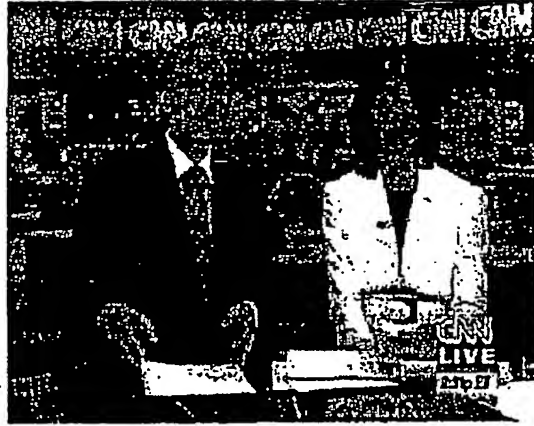


图 10A



图 10B

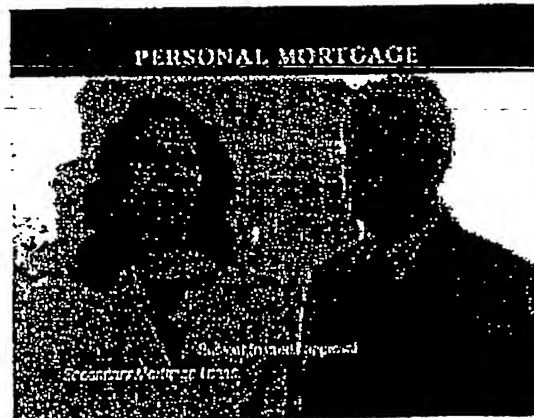


图 10C



图 10D



图 10E



图 10F

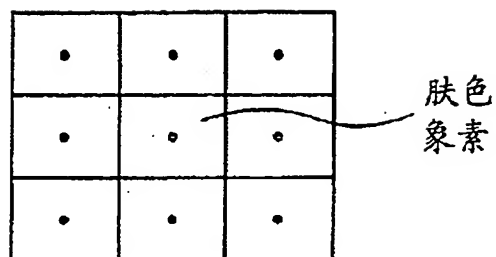


图 11

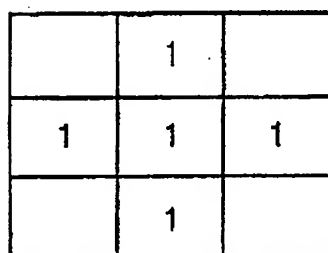


图 12A

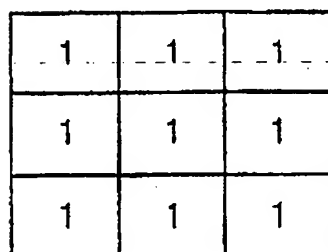


图 12B



图 13A



图 13B



图 14

00-10-23

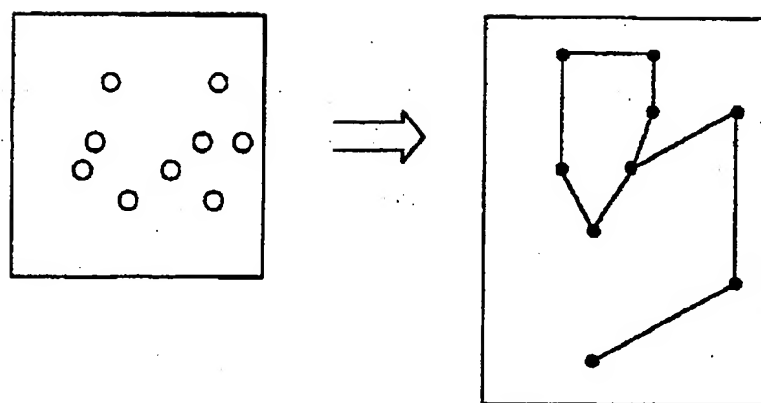


图 15



图 16A



图 16B



图 16C



图 16D

00-10-23

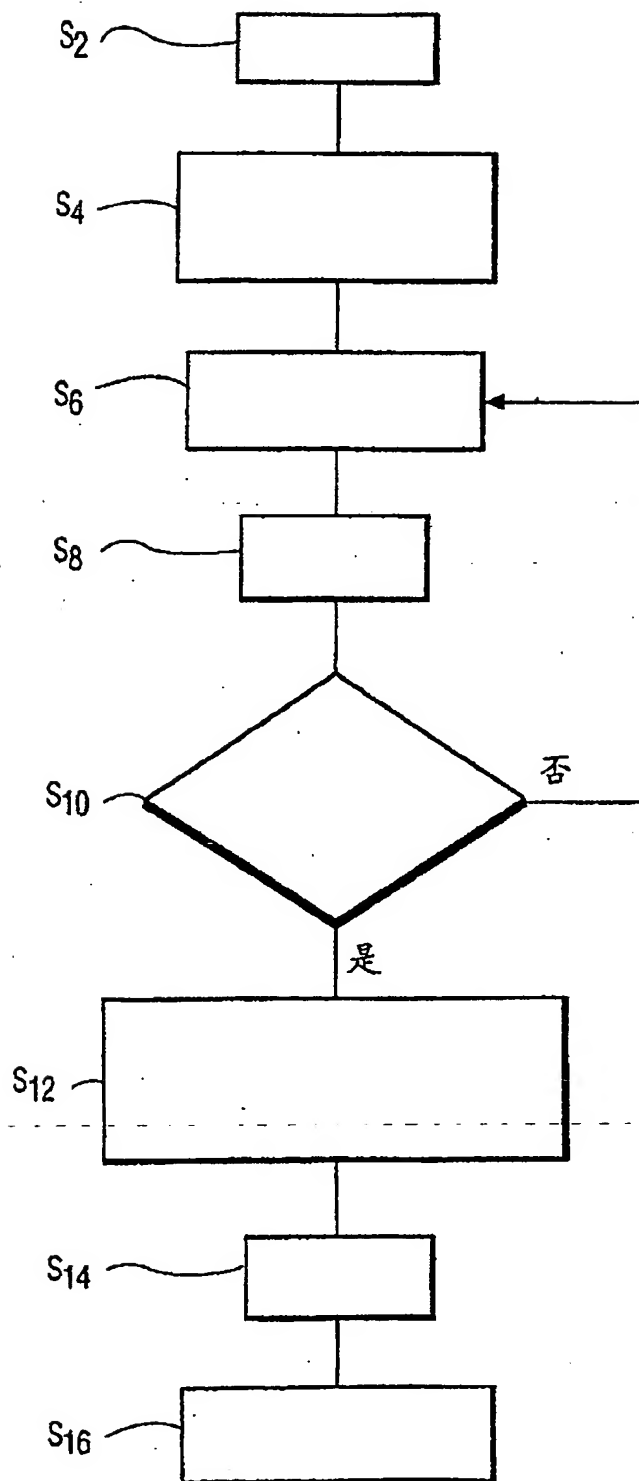


图 17

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.